network\_analysis with qgraph

Gabriel Burcea

07/08/2020

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.1 ✓ purrr 0.3.4  
## ✓ tibble 3.0.3 ✓ dplyr 1.0.0  
## ✓ tidyr 1.1.0 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.5.0

## ── Conflicts ────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(cluster)  
library(plotly)

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(fpc)  
library(dendextend)

##   
## ---------------------  
## Welcome to dendextend version 1.13.4  
## Type citation('dendextend') for how to cite the package.  
##   
## Type browseVignettes(package = 'dendextend') for the package vignette.  
## The github page is: https://github.com/talgalili/dendextend/  
##   
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues  
## Or contact: <tal.galili@gmail.com>  
##   
## To suppress this message use: suppressPackageStartupMessages(library(dendextend))  
## ---------------------

##   
## Attaching package: 'dendextend'

## The following object is masked from 'package:stats':  
##   
## cutree

library(factoextra)

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

library(FactoMineR)  
library(NbClust)  
library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library(DMwR)

## Loading required package: grid

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

library(ggraph)

data\_cluster <- read\_csv("/Users/gabrielburcea/rprojects/stats\_data\_whole/data\_categ\_no\_sev.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_character(),  
## age = col\_double(),  
## date\_completed = col\_date(format = ""),  
## how\_unwell = col\_double(),  
## number\_days\_symptoms = col\_double(),  
## number\_morbidities = col\_double()  
## )

## See spec(...) for full column specifications.

## Warning: 9 parsing failures.  
## row col expected actual file  
## 2910 how\_unwell a double How Unwell '/Users/gabrielburcea/rprojects/stats\_data\_whole/data\_categ\_no\_sev.csv'  
## 10025 number\_days\_symptoms no trailing characters 0 '/Users/gabrielburcea/rprojects/stats\_data\_whole/data\_categ\_no\_sev.csv'  
## 19970 number\_days\_symptoms no trailing characters 9 '/Users/gabrielburcea/rprojects/stats\_data\_whole/data\_categ\_no\_sev.csv'  
## 22410 number\_days\_symptoms a double Plus de 21 '/Users/gabrielburcea/rprojects/stats\_data\_whole/data\_categ\_no\_sev.csv'  
## 31280 number\_days\_symptoms no trailing characters ?? ???? ?? '/Users/gabrielburcea/rprojects/stats\_data\_whole/data\_categ\_no\_sev.csv'  
## ..... .................... ...................... ........... .......................................................................  
## See problems(...) for more details.

level\_key\_temperature <- c("Yes" = "37.5-38",   
 "Yes" = "38.1-39",   
 "Yes" = "38.2-39",  
 "Yes" = "39.1-41")  
  
  
  
data\_select <- data\_cluster %>%  
 dplyr::select(ID, covid\_tested, chills, cough, diarrhoea, fatigue, headache, loss\_smell\_taste, muscle\_ache,   
 nasal\_congestion, nausea\_vomiting, shortness\_breath, sore\_throat, sputum, temperature, asthma, diabetes\_type\_one,  
 diabetes\_type\_two, obesity, hypertension, heart\_disease, lung\_condition, liver\_disease, kidney\_disease) %>%  
 dplyr::filter(covid\_tested != "none")   
  
covid\_tested\_levels <- c("positive" = "showing symptoms")  
  
data\_transf <- data\_select %>%   
 dplyr::mutate(covid\_tested = forcats::fct\_recode(covid\_tested, !!!covid\_tested\_levels),   
 temperature = forcats::fct\_recode(temperature, !!!level\_key\_temperature)) %>%  
 dplyr::filter(covid\_tested != "none")

## Warning: Unknown levels in `f`: 38.2-39

data\_transf$temperature <- as.character(data\_transf$temperature)  
  
data\_transf$ID <- NULL  
data\_transf$covid\_tested <- NULL  
  
level\_key\_comorbidities <-  
 c("kidney disease" = "kidney\_disease",  
 "lung condition" = "lung\_condition",  
 "diabetes type one" = "diabetes\_type\_one",  
 "diabetes type two" = "diabetes\_type\_two",  
 "liver disease" = "liver\_disease",  
 "heart disease" = "heart\_disease")  
  
  
data\_pi<- data\_transf %>%  
 tidyr::replace\_na(list(nausea\_vomiting = 0)) %>%  
 dplyr::rename('loss of smell and taste' = loss\_smell\_taste, 'muscle ache' = muscle\_ache, 'nasal congestion'= nasal\_congestion,  
 'nausea and vomiting' = nausea\_vomiting, 'shortness of breath' = shortness\_breath, 'sore throat' = sore\_throat,   
 "kidney disease" = kidney\_disease, "diabetes type one" = diabetes\_type\_one, "diabetes type two" = diabetes\_type\_two,  
 "liver disease" = liver\_disease, "heart disease" = heart\_disease)  
  
  
# data\_piv <- data\_transf %>%  
# pivot\_longer(cols = 14:22,  
# names\_to = "Comorbidities",  
# values\_to = "Bolean") %>%  
# dplyr::filter(Bolean == "Yes") %>%  
# data\_piv <- data\_piv %>%  
# pivot\_wider(names\_from = Comorbidities, values\_from = Freq) %>%  
# pivot\_longer(cols = 3:15,  
# names\_to = "Symptoms",  
# values\_to = "Yes\_No") %>%  
# dplyr::group\_by(Comorbidities, Symptoms, Yes\_No) %>%  
# dplyr::summarise(Count = n()) %>%  
# dplyr::mutate(Freq = Count / sum(Count)\*100) %>%  
# dplyr::filter(Yes\_No == "Yes")  
  
  
# data\_piv$Bolean <- NULL  
  
#   
  
# data\_piv$Comorbidities <- as.character(data\_piv$Comorbidities)  
  
  
  
# scale data

This is unregularised partial correlation network Threshold argument - removes the edges that are not significant. If I pass any threshold a lot of edges dissapear

library(qgraph)

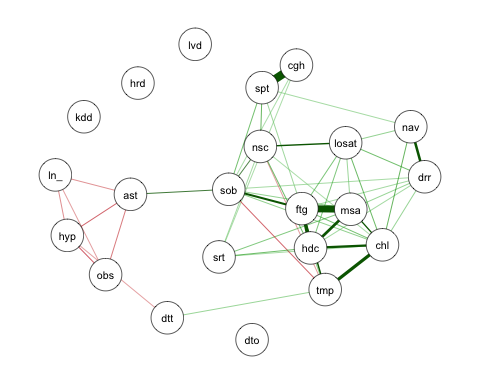
## Registered S3 methods overwritten by 'BDgraph':  
## method from  
## plot.sim lava  
## print.sim lava

## Registered S3 methods overwritten by 'huge':  
## method from   
## plot.roc pROC   
## plot.sim BDgraph  
## print.roc pROC   
## print.sim BDgraph

##   
## Attaching package: 'qgraph'

## The following object is masked from 'package:ggraph':  
##   
## qgraph

cor\_mat <- cor\_auto(data\_pi)  
  
view(round(cor\_mat,2))  
graph\_pcor <- qgraph(cor\_mat, graph = "pcor", layout = "spring", threshold = "bonferroni", sampleSize = nrow(data\_pi), alpha = 0.01) # t

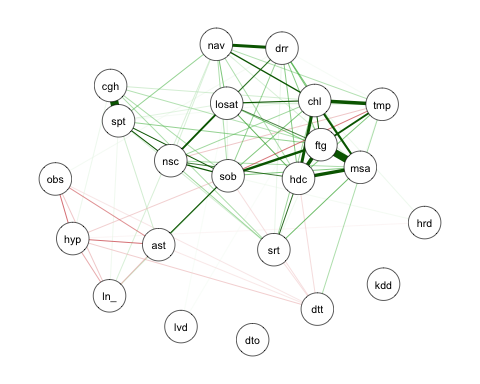


# threshold Inadditiontoanumericvaluetoomitedgesthisargumentcanalsobeassignedastringto omit insignficant edges. Note that this REMOVES edges from the network (which influences centrality measures and the spring layout). Can be "sig" to compute significance without correction for multiple testing, "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr" or "none" which are used directly in the adjust argument in corr.p of the psych package (Revelle, 2014). In addition, this argument can be assigned "locfdr" in which edges are set to zero if the local FDR is below FDRcutoff. fdrtool from the fdrtool package (Klaus and Strimmer, 2014) is used to compute these measures, which is used inside FDRnetwork.

Estimating a partial correlation network using LASSO regularization and EBIC model selection can be done by setting graph = “glasso”. The tuning argument sets the EBIC hyperparameter. Set between 0 (more connections but also more spurious connections) and 0.5 (more parsimony, but also missing more connections):

graph\_lasso <- qgraph(cor\_mat, graph = "glasso", layout = "spring", tuning = 0.25,   
 sampleSize = nrow(data\_pi))

## Warning in EBICglassoCore(S = S, n = n, gamma = gamma, penalize.diagonal =  
## penalize.diagonal, : A dense regularized network was selected (lambda < 0.1 \*  
## lambda.max). Recent work indicates a possible drop in specificity. Interpret the  
## presence of the smallest edges with care. Setting threshold = TRUE will enforce  
## higher specificity, at the cost of sensitivity.



Centrality analysis

centRes <- centrality(graph\_lasso)  
  
# Node strenght (degree):  
  
centRes$OutDegree

## chl cgh drr ftg hdc losat   
## 0.949934865 0.469085440 0.529115887 1.049956223 0.888986457 0.579366665   
## msa nsc nav sob srt spt   
## 0.888002724 0.540940281 0.486190164 0.732794942 0.290627790 0.664188733   
## tmp ast dto dtt obs hyp   
## 0.504691966 0.238703219 0.000000000 0.100227235 0.133253523 0.191298894   
## hrd ln\_ lvd kdd   
## 0.013459255 0.114077557 0.004874035 0.000000000

# Closeness:  
  
centRes$Closeness

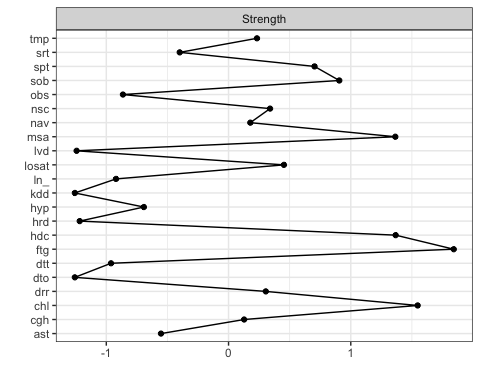
## chl cgh drr ftg hdc losat msa nsc nav sob srt spt tmp   
## 0 0 0 0 0 0 0 0 0 0 0 0 0   
## ast dto dtt obs hyp hrd ln\_ lvd kdd   
## 0 0 0 0 0 0 0 0 0

# Betweeenness:   
centRes$Betweenness

## chl cgh drr ftg hdc losat msa nsc nav sob srt spt tmp   
## 80 2 0 100 40 40 44 32 12 138 0 34 30   
## ast dto dtt obs hyp hrd ln\_ lvd kdd   
## 60 0 0 0 6 0 0 0 0

centralityPlot(graph\_lasso)

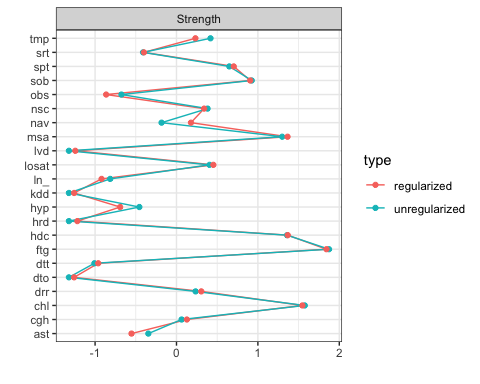
## Note: z-scores are shown on x-axis rather than raw centrality indices.



compare different networks

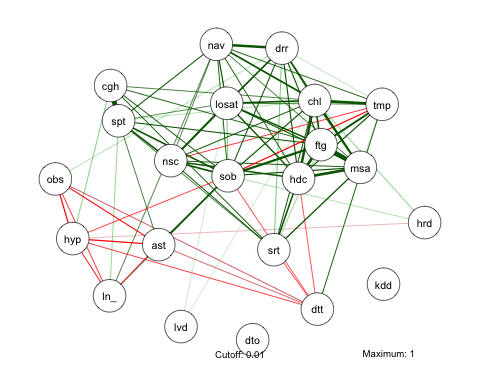
centralityPlot(GGM = list(unregularized = graph\_pcor, regularized = graph\_lasso))

## Note: z-scores are shown on x-axis rather than raw centrality indices.

 to make edges in graphs comparable in qgraph, the cut, minimum and maximum arguments need to be set to the same values. details = TRUE

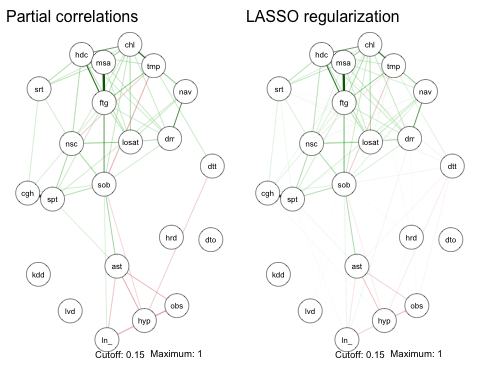
qgraph(cor\_mat, graph = "glasso", layout = "spring", tuning = 0.25,   
 sampleSize = nrow(data\_pi), minimum = 0, cut = 0.01, maximum = 1, details = TRUE,   
 esize = 20) # the more I increase the cut the less edges are present - I have tried with 0.15 , 0.10

## Warning in EBICglassoCore(S = S, n = n, gamma = gamma, penalize.diagonal =  
## penalize.diagonal, : A dense regularized network was selected (lambda < 0.1 \*  
## lambda.max). Recent work indicates a possible drop in specificity. Interpret the  
## presence of the smallest edges with care. Setting threshold = TRUE will enforce  
## higher specificity, at the cost of sensitivity.

 Comparable layouts

Layout <- averageLayout(graph\_pcor,graph\_lasso)  
layout(t(1:2))  
qgraph(cor\_mat, graph = "pcor", layout = Layout,threshold = "bonferroni",  
 sampleSize = nrow(data\_pi), minimum = 0,   
 cut = 0.15, maximum = 1, details = TRUE,  
 esize = 20, title = "Partial correlations") #   
  
qgraph(cor\_mat, graph = "glasso", layout = Layout, tuning = 0.25,  
 sampleSize = nrow(data\_pi), minimum = 0,  
 cut = 0.15, maximum = 1, details = TRUE,  
 esize = 20, title = "LASSO regularization")

## Warning in EBICglassoCore(S = S, n = n, gamma = gamma, penalize.diagonal =  
## penalize.diagonal, : A dense regularized network was selected (lambda < 0.1 \*  
## lambda.max). Recent work indicates a possible drop in specificity. Interpret the  
## presence of the smallest edges with care. Setting threshold = TRUE will enforce  
## higher specificity, at the cost of sensitivity.



qgraph(cor\_mat, graph = "glasso", layout = "spring", tuning = 0.25,  
 sampleSize = nrow(data\_pi), legend.cex = 0.2, vsize = 5,  
 esize = 15, pastel = TRUE, posCol = "#003399",  
 negCol = "#FF9933", borders = FALSE, vTrans = 200,  
 details = TRUE)

## Warning in qgraph(cor\_mat, graph = "glasso", layout = "spring", tuning = 0.25, :  
## The 'pastel' argument is deprecated, please use palette = 'pastel' instead.

## Warning in EBICglassoCore(S = S, n = n, gamma = gamma, penalize.diagonal =  
## penalize.diagonal, : A dense regularized network was selected (lambda < 0.1 \*  
## lambda.max). Recent work indicates a possible drop in specificity. Interpret the  
## presence of the smallest edges with care. Setting threshold = TRUE will enforce  
## higher specificity, at the cost of sensitivity.

